Volume 301, number 1, 83–88 FEBS 10921 April 1992

Cloning and sequence analysis of the phenylalanyl-tRNA synthetase genes (*pheST*) from *Thermus thermophilus*

Barbara Keller, Peter Kast and Hauke Hennecke

Mikrobiologisches Institut, Eidgenössische Technische Hochschule, Schmelzbergstrasse 7, CH-8092 Zürich, Switzerland

Received 28 February 1992

While crystals suitable for X-ray diffraction analyses are available of phenylalanyl-tRNA synthetase (PheRS) from the thermophilic bacterium *Thermus thermophilus*, neither the primary structure of its constituent α and β subunits nor the nucleotide sequence of the corresponding *pheS* and *pheT* genes were known. Using specific oligonucleotides of conserved *pheS* regions that were adapted to the *T. thermophilus* codon usage, we identified, cloned and subsequently sequenced the *pheST* genes of this bacterium. The sequences reported here will greatly aid in the three-dimensional structure determination of *T. thermophilus* PheRS, a heterotetrameric ($\alpha_2\beta_2$), class II aminoacyl-tRNA synthetase.

Class II aminoacyl-tRNA synthetase; Gene cloning; Nucleotide sequence; Phenylalanyl-tRNA synthetase; *pheS* and *pheT* gene

1. INTRODUCTION

The highly specific selection of amino acid and tRNA substrates by the 20 aminoacyl-tRNA synthetases is an intriguing problem in protein biosynthesis. Significant progress in the interpretation of structure/function and evolutionary relationships has been made recently with the discovery of a partition of these enzymes into two classes based on mutually exclusive sets of sequence motifs [1–5] and the three-dimensional structure determination of five aminoacyl-tRNA synthetases, three of class I [6–8] and two of class II [9,10].

The three-dimensional structure determination of a prokaryotic PheRS, a class II enzyme, would be of particular interest for at least two reasons. (i) Together with glycyl-tRNA synthetase, bacterial PheRSs belong to the two high-molecular weight aminoacyl-tRNA synthetases having the unusual $\alpha_2\beta_2$ subunit structure. Although there is some evidence that the β subunit (M_r~90,000) might participate in $tRNA^{Phe}$ binding [11,12], it is intriguing to note that all of the conserved sequence motifs characteristic for class II tRNA synthetases are located in the smaller α subunit (M_r~40,000). Recently, the surprising discovery was made that mitochondrial PheRS from the yeast *Saccharomyces cerevisiae* is active as a monomer; this protein corresponds to the bacterial α subunit which, however, has recruited an additional, C-terminal 100-amino acid extension that shares significant sequence similarity with a C-terminal domain present in bacterial PheRS β subunits [13]. (ii) In contrast to the nine other members of class II tRNA synthetases, PheRS makes an exception in that it catalyzes the initial aminoacylation of the 2′ hydroxyl of the terminal adenosine of tRNA rather than the 3′ hydroxyl [14,15]. The general distinction between 2′OH and 3′OH aminoacylation by class I and class II tRNA synthetases, respectively, has given rise to – among other considerations – the hypothesis of two evolutionarily different tRNA synthetase ancestors [1]. Furthermore, this difference may reflect different topological modes of tRNA binding to the enzymes in order to facilitate a correct stereospecific positioning of one of the terminal ribose hydroxyls near the active site [10]. In this context, it would be desirable to unravel the structure of PheRS-$tRNA^{Phe}$ co-crystals.

While PheRS has been purified from various organisms, crystallization has so far been achieved only in the case of PheRS from the extreme thermophile *Thermus thermophilus* strain HB8 [16,17]. The crystals obtained are of a quality suitable for X-ray diffraction analysis down to 3.5 Å resolution. Further detailed structural investigations, however, will be greatly hampered by the fact that no primary structure information is available on the α and β subunits [16,18] of *T. thermophilus* PheRS. We have therefore decided to clone the respective structural genes (*pheS* for the α subunit and *pheT* for the β subunit) and determine their nucleotide sequences. The approach we used was to search for regions with highest similarity in three PheRS small subunit sequences that were published at the beginning of this work, i.e., those of *Escherichia coli* [19,20] and *S. cerevisiae* cytoplasmic [21] and mitochondrial [22] PheRS. Specific oligonucleotides corresponding to these homologous protein stretches were then designed

Abbreviations: bp, basepair(s); kb, kilobasepair(s); PheRS, phenylalanyl-tRNA synthetase.

Correspondence address: H. Hennecke, Mikrobiologisches Institut, Eidgenössische Technische Hochschule, Schmelzbergstrasse 7, CH-8092 Zürich, Switzerland. Fax: (41) (1) 262 0647.

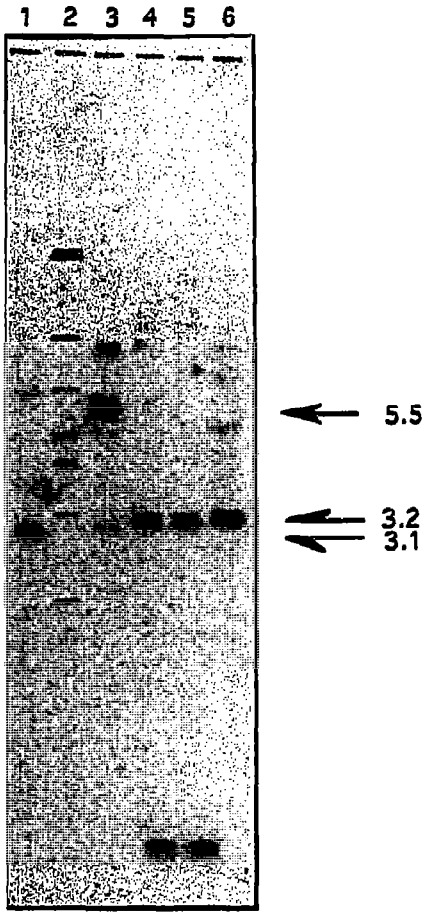


Fig. 1 Autoradiogram of the Southern blot hybridization of *T. thermophilus* chromosomal DNA with oligonucleotide TTH1. Arrows on the right indicate the sizes (in kb) of the main hybridizing fragments. In lane 2 the positions of phage Lambda DNA fragments (digested with *Eco*RI/*Hin*dIII and *Hin*dIII) are marked. In lane 1 *T. thermophilus* DNA is digested with *Sph*I/*Hin*dIII, in lane 3 with *Sph*I/ *Xba*I, in lane 4 with *Xba*I/*Kpn*I, in lane 5 with *Kpn*I, and in lane 6 with *Kpn*I/ *Hin*dIII.

by taking into consideration the codon usage in *T. thermophilus* genes whose sequences were known. The use of oligonucleotides as hybridization probes led to the successful identification and cloning first of the *T. thermophilus pheS* gene and subsequently also of the *pheT* gene which turned out to be located immediately adjacent (downstream) of *pheS*.

Fig. 3 Nucleotide and deduced amino acid sequence of the *T. thermophilus pheST* region as indicated in Fig. 2. The *pheS* gene extends from positions 327 to 1379 and *pheT* from 1376 to 3733. Potential ribosome binding sites in front of the two genes are overscored.

→

2. MATERIALS AND METHODS

2.1. *Bacterial strains, vectors and plasmids*

Thermus thermophilus strain HB8 [23] was a gift from L. Reshetnikova, Moscow. *E. coli* strain JM109 and JM101 [24] were the hosts for cloning and single strand DNA isolation, respectively. Vectors used were pUC19 [24] for cloning, pBluescript KS(+) [25] for construction of deletions, and M13mp18 and M13mp19 [26] for single strand DNA sequencing. Plasmids pKST-KB1 and pKST-KB2 were constructed with pBluescript and contain a 3.2 kb *Kpn*I fragment in both orientations, whereas pKST-Ba19U contains a 4.6 kb *Bam*H1 fragment ligated to *Bam*H1-digested pUC19; both fragments are from *T. thermophilus* strain HB8 [23] chromosomal DNA.

2.2. *Cloning procedures*

Three oligonucleotides were designed for Southern blot analysis in order to find the *pheS* gene of *T. thermophilus*. An alignment of the amino acid sequences of the PheRS α subunits from *E. coli* [19,20], and the *S. cerevisiae* cytoplasmic [21] and mitochondrial [22] enzymes was made, and three particularly well-conserved oligopeptides were selected: NFDAL (positions 139–143 in the *E. coli* sequence), YFPFTEP (247–253) and GKWLE (264–268), to which corresponding oligonucleotides TTH1 5′AACTTTGACIC$^{G}_{C}$CTG3′ (I=Inosine), TTH4 5′TACTTCCC$^{G}_{C}$TIIAC$^{G}_{C}$GAGCC3′ and TTH5 5′CTC$^{G}_{C}$AGCCACT$^{TG}_{CC}$CC3′ (non-coding strand) were designed by adapting the *E. coli* nucleotide sequence to the *T. thermophilus* codon usage. The *T. thermophilus* codon usage chart was calculated from all of the *T. thermophilus* gene sequences found in the databank in the year 1989. Furthermore, a 506 bp *Bst*EII/*Bst*BI fragment from the *E. coli pheS* gene (positions 532–1037, [19]) was used as additional hybridization probe. Hybridizations were performed under non-stringent conditions [27].

Restriction enzyme digested chromosomal DNA from *T. thermophilus* HB8 was run on a 0.8% agarose gel and fragments of appropriate size (according to hybridization results) were excised and ligated to digested pBluescript or pUC19 vectors. The resulting plasmids were

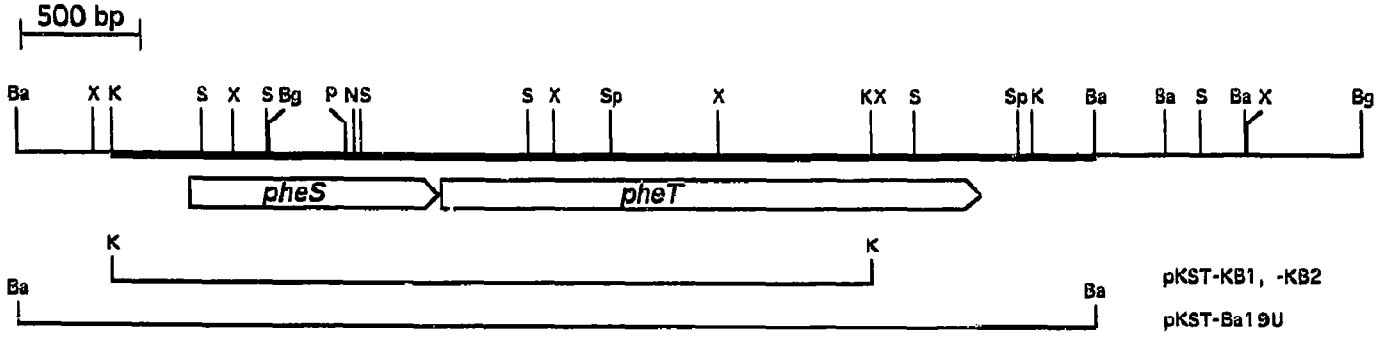
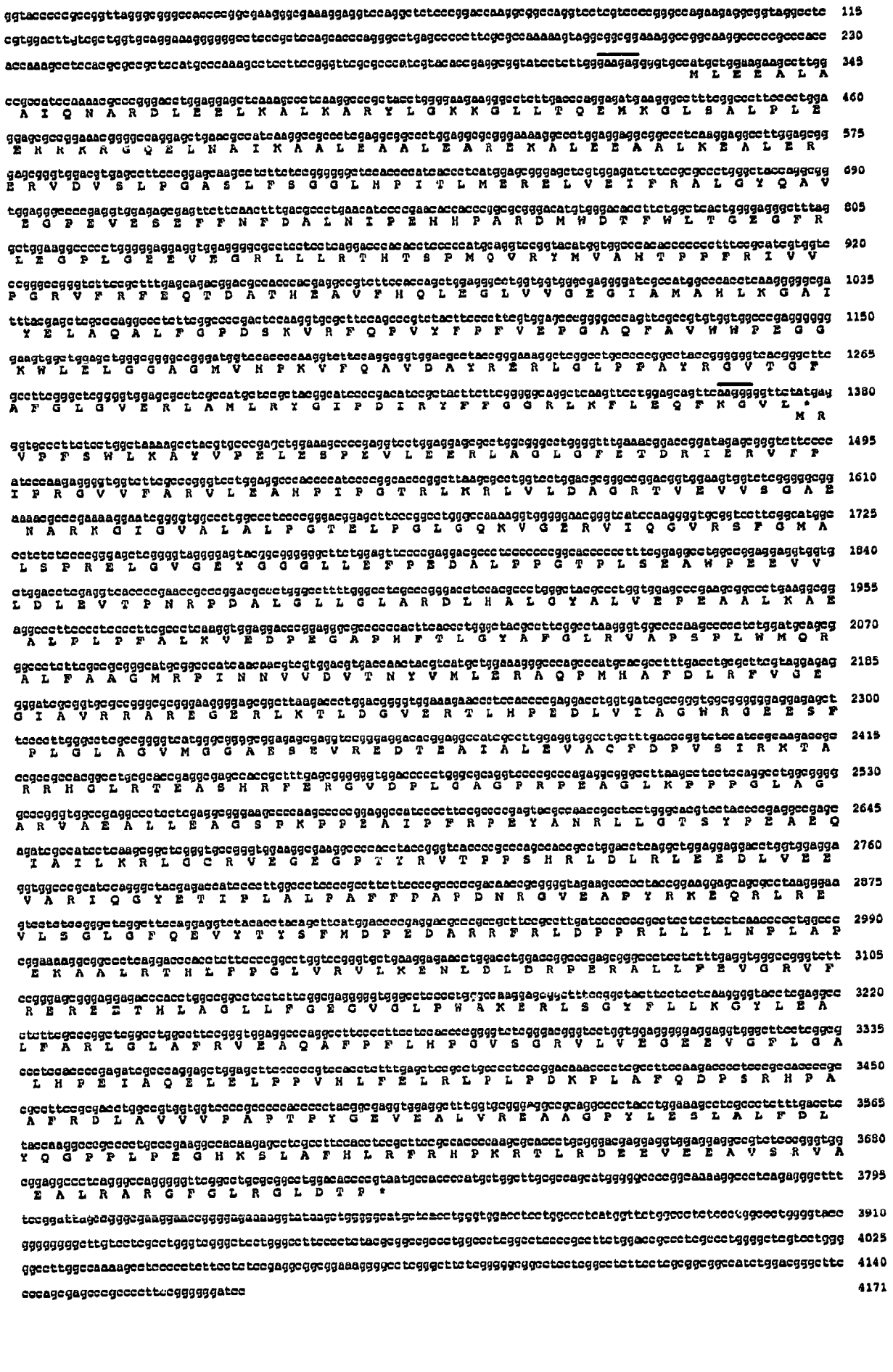


Fig. 2 Restriction map of the genomic DNA region from *T. thermophilus* containing the *pheS* and *pheT* genes. The sequenced part is indicated by a bold line. Plasmid inserts mentioned in the text are shown below the map. Abbreviations: Ba, *Bam*H1; Bg, *Bgl*II; K, *Kpn*I; P, *Pvu*II; N, *Nco*I; S, *Sac*I; Sp, *Sph*I; X, *Xho*I.



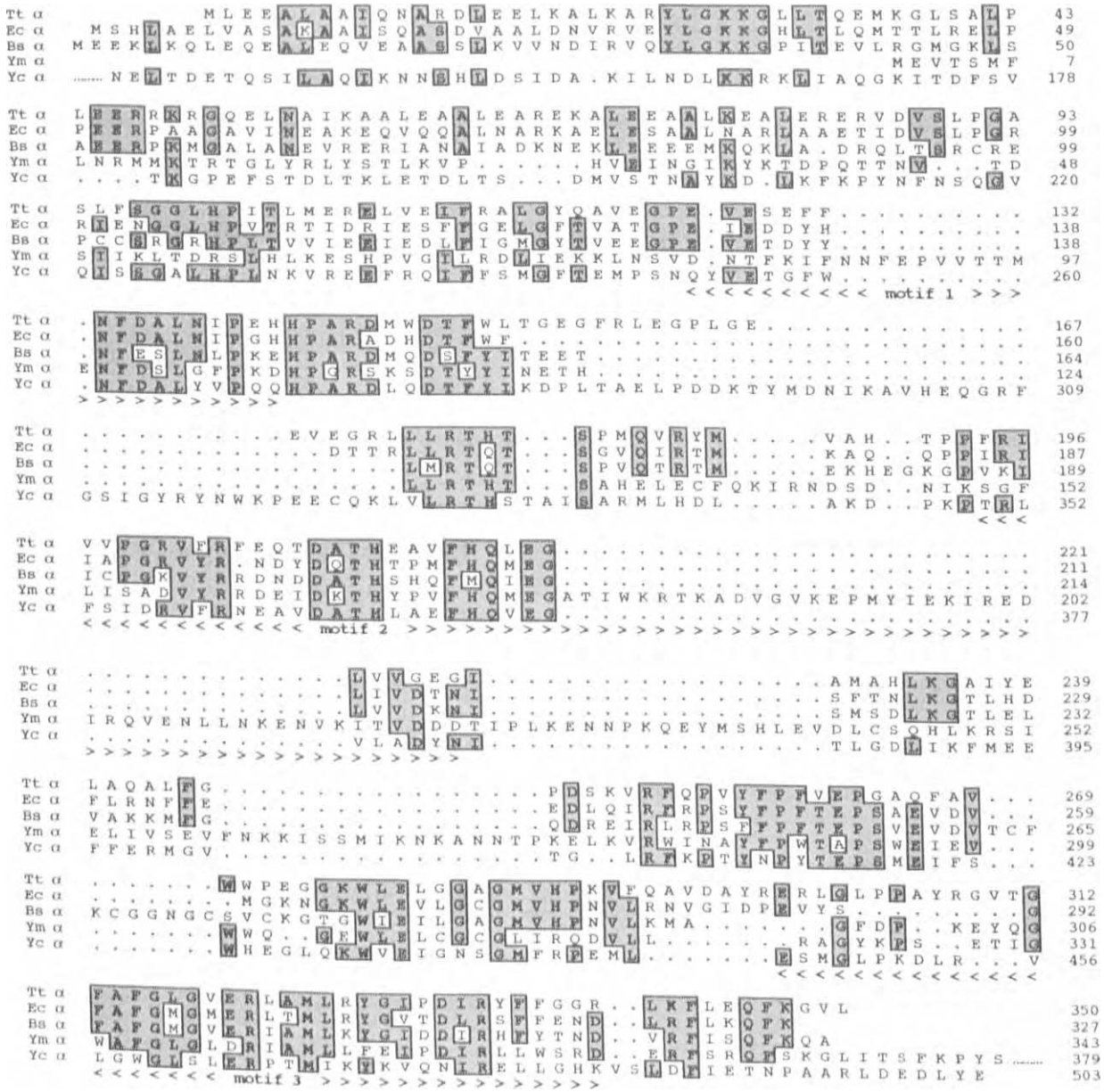


Fig. 4 Alignment of known PheRS α subunit amino acid sequences. Identical residues are boxed. The shown PheRS α subunits are from *T. thermophilus* (this work) *E. coli* [19,20], *B. subtilis* [30], *S. cerevisiae* mitochondria [22] and cytoplasm [21] (from top to bottom); of the last two sequences only the relevant parts are shown.

transformed into *E. coli* cells and screened by colony hybridization with oligonucleotide TTH1.

2.3. *Sequencing strategy*

A set of 12 deletions of the plasmids pKST-KB1 and pKST-KB2 was constructed using the ExoIII/mung bean nuclease method [28]; the plasmids were then used for double strand sequencing applying the dideoxy chain termination method [29] after a denaturation step with NaOH. To obtain the complete overlapping sequence on both strands, suitable subclones were made from plasmids pKST-KB1, pKST-KB2, and pKST-Ba19U. All sequencing reactions were run on an automated DNA Sequencer model 373A (Applied Biosystems, Foster City, CA) and carried out following the protocols given by the manufacturer.

3. RESULTS AND DISCUSSION

In order to find a suitable fragment for cloning of the *pheS* gene and possibly also the *pheT* gene from *T. thermophilus*, Southern blot analyses were performed as detailed in section 2.2. Fig. 1 shows an autoradiogram of the hybridization with oligonucleotide TTH1. The same major bands seen here were also detected with the other hybridization probes (TTH4, TTH5, 506 bp *E. coli pheS* fragment), and it was therefore assumed that they contained the desired gene.

Plasmids pKST-KB1 and pKST-KB2 were con-

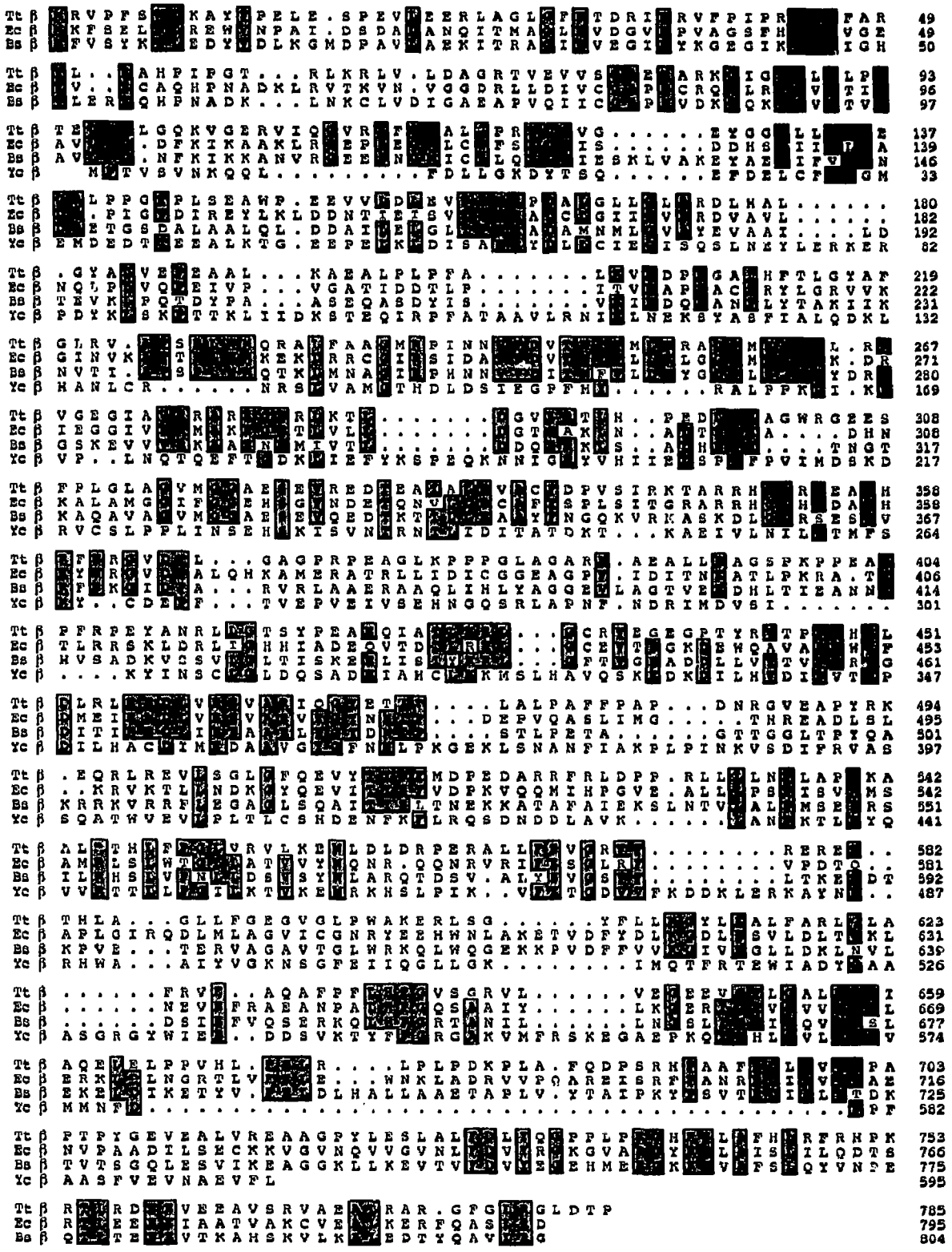


Fig. 5: Alignment of known PheRS β subunit amino acid sequences. The shown sequences are from *T. thermophilus* (this work), *E. coli* [32], *B. subtilis* [30] and *S. cerevisiae* cytoplasm [21] (from top to bottom).

structed by cloning a 3.2 kb *Kpn*I fragment that was detected in the hybridizations. The same fragment was then used as a homologous hybridization probe in order to find additional DNA fragments, resulting in the cloning of a 4.6 kb *Bam*HI fragment, to give plasmid pKST-Ba19U. A restriction map from the region was created

(Fig. 2) and the 4.15 kb stretch indicated in Fig. 2 was sequenced. The complete nucleotide sequence is shown in Fig. 3. We found that the sequenced region indeed contained the *pheS* gene, as well as the *pheT* gene, which we could locate downstream of *pheS*. The *pheS* gene starts at position 327 and ends at position 1379 and is thus 1,053 nucleotides long. The *pheT* gene starts at position 1,376 of the shown sequence and, with a length of 2,358 bp, terminates at position 3,733 (Fig. 3). The *pheT* gene thus overlaps by 4 nucleotides with the *pheS* gene, indicating that the genes are most probably transcribed in an operon. The situation in *E. coli* is slightly different: although the genes are also organized in an operon, they are separated by 14 nucleotides [19]. The G+C-content of the two *T. thermophilus* coding regions is 70.5% and, hence, almost the same as the overall G+C-content of that bacterium (69% [23]). Moreover, the codon usage is typically biased, e.g., of the two codons for threonine the one that ends with C is exclusively used.

Figures 4 and 5 show an alignment of the deduced amino acid sequences of the known PheRS α and β subunits, respectively, with the predicted sequences of the *T. thermophilus pheS* and *pheT* gene products. In general, the α subunits display a greater homology; the identity to the *T. thermophilus* α subunit amino acid sequence ranges from 32.5% for the yeast mitochondrial enzyme to 48.5% for the *E. coli* enzyme. The β subunit gives a much lower score with identities between 21.4% for the yeast cytoplasmic PheRS and 36.4% for the *E. coli* PheRS.

In Fig. 4 the three sequence motifs that define class II aminoacyl-tRNA synthetases [1] are indicated. Motif 1 is also shown, even though it is believed to be involved in the formation of the subunit interface of the class II synthetases having an α_2 subunit structure [2]. In the case of prokaryotic phenylalanyl-tRNA synthetases which have an $\alpha_2\beta_2$ subunit structure the role of motif 1 remains unclear. Motifs 2 and 3 are supposed to form a cavity where aminoacylation takes place, i.e. the active site of the enzyme [9,10,20,31]. These motifs are well conserved in *T. thermophilus*. It is interesting to note that the peptide GFAFG (positions 312–316 in *T. thermophilus*) in motif 3 is totally homologous to the corresponding *E. coli* sequence. In previous work, this region was suggested to participate in the binding of the substrate phenylalanine [20].

Another region of interest is the C-terminus of the β-subunits (Fig. 5). The stretch from positions 685–780 (numbering refers to *T. thermophilus*) is well conserved in the bacterial enzymes, and is found as an extension at the C-terminus of the yeast mitochondrial PheRS, which otherwise is an α subunit homologue [13]. It is speculated that this extension confers upon the yeast mitochondrial PheRS an independence of the β subunit.

REFERENCES

[1] Eriani, G., Delarue, M., Poch, O., Gangloff, J. and Moras, D. (1990) Nature 347, 203–206.
[2] Cusack, S., Härtlein, M. and Leberman, R. (1991) Nucleic Acids Res. 19, 3489–3498.
[3] Nagel, G.M. and Doolittle, R.F. (1991) Proc. Natl. Acad. Sci. USA 88, 8121–8125.
[4] Burbaum, J.J. and Schimmel, P. (1991) J. Biol. Chem. 266, 16965–16968.
[5] Mirande, M. (1991) Progr. Nucleic Acids Res. Mol. Biol. 40, 95–142.
[6] Brunie, S., Zelwer, C. and Risler, J.-L. (1990) J. Mol. Biol. 216, 411–424.
[7] Brick, P., Bhat, T.N. and Blow, D.M. (1989) J. Mol. Biol. 208, 83–98.
[8] Rould, M.A., Perona, J.J., Söll, D. and Steitz, T.A. (1989) Science 246, 1135–1142.
[9] Cusack, S., Berthet-Colominas, C., Härtlein, M., Nassar, N. and Leberman, R. (1990) Nature 347, 249–255.
[10] Ruff, M., Krishnaswamy, S., Boeglin, M., Poterszman, A., Mitschler, A., Podjarny, A., Rees, B., Thierry, J.C. and Moras, D. (1991) Science 252, 1682–1689.
[11] Khodyreva, S.N., Moor, N.A., Ankilova, V.N. and Lavrik, O.I. (1985) Biochim. Biophys. Acta 830, 206–212.
[12] Hountondji, C., Schmitter, J.-M., Beauvallet, C. and Blanquet, S. (1987) Biochemistry 26, 5433–5439.
[13] Sanni, A., Walter, P., Boulanger, Y., Ebel, J.-P. and Fasiolo, F. (1991) Proc. Natl. Acad. Sci. USA 88, 8387–8391.
[14] Sprinzl, M. and Cramer, F. (1975) Proc. Natl. Acad. Sci. USA 72, 3049–3053.
[15] Fraser, T.H. and Rich, A. (1975) Proc. Natl. Acad. Sci. USA 72, 3044–3048.
[16] Ankilova, V.N., Reshetnikova, L.S., Chernaya, M.M. and Lavrik, O.I. (1988) FEBS Lett. 227, 9–13.
[17] Chernaya, M.M., Korolev, S.V., Reshetnikova, L.S. and Safro, M.G. (1987) J. Mol. Biol. 198, 555–556.
[18] Bobkova, E.V., Mashanov-Golikov, A.V., Wolfson, A., Ankilova, V.N. and Lavrik, O.I. (1991) FEBS Lett. 290, 95–98.
[19] Fayat, G., Mayaux, J.-F., Sacerdot, C., Fromant, M., Springer, M., Grunberg-Manago, M. and Blanquet, S. (1983) J. Mol. Biol. 171, 239–261.
[20] Kast, P. and Hennecke, H. (1991) J. Mol. Biol. 222, 99–124.
[21] Sanni, A., Mirande, M., Ebel, J.-P., Boulanger, Y., Waller, J.-P. and Fasiolo, F. (1988) J. Biol. Chem. 263, 15407–15415.
[22] Koerner, T.J., Myers, A.M., Lee, S. and Tzagoloff, A. (1987) J. Biol. Chem. 262, 3690–3696.
[23] Oshima, T. and Imahori, K. (1974) Int. J. Syst. Bacteriol. 24, 102–112.
[24] Yanisch-Perron, C., Vieira, J. and Messing J. (1985) Gene 33, 103–119.
[25] Short, J.M., Fernandez, J.M., Sorge, J.A. and Huse, W.D. (1988) Nucleic Acids Res. 16, 7583–7600.
[26] Norrander, J., Kempe, T. and Messing, J. (1983) Gene 26, 101–106.
[27] Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning. A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
[28] Henikoff, S. (1984) Gene 28, 351–359.
[29] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[30] Brakhage, A.A., Wozny, M. and Putzer, H. (1990) Biochimie 72, 725–734 (Erratum (1991): Biochimie 73, 127).
[31] Kast, P., Wehrli, C. and Hennecke, H. (1991) FEBS Lett. 293, 160–163.
[32] Miller, H.I. (1986) EMBL Sequence Database Accession Number K02844.